

Novas Medidas de Educação e de Desigualdade Educacional para a Primeira Metade do Século XX no Brasil

Bruno Komatsu¹

Naercio Menezes-Filho²

Pedro Augusto Costa Oliveira³

Leonardo Teixeira Viotti⁴

Resumo

Este artigo descreve a trajetória do nível e da desigualdade educacional no Brasil desde a primeira metade do século XX. Combinamos diversas fontes de dados históricos, tais como os relatórios do Ministério de Negócios do Império, os Anuários Estatísticos do Brasil e os Censos Demográficos para construir novas medidas de escolaridade e calcular índices de Gini educacional entre 1900 e 2000 para cada região e para o Brasil como um todo. Nossos resultados mostram que entre 1900 e 1930, a proporção de pessoas com ensino primário completo na população permaneceu em torno de 5%, ao passo que a parcela com ensino secundário completo esteve sempre abaixo de 1% e que somente 0,3% tinham ensino superior completo. Assim, a desigualdade educacional permaneceu constante até 1920, declinou lentamente entre 1920 e 1950 e mais rapidamente somente a partir de então.

Palavras-Chave

Educação. Desigualdade. Índice de Gini.

Abstract

This paper describes the trajectory of educational level and inequality in Brazil in the first half of the 20th century. We combine various sources of historical data, such as the reports of the Ministry of Business of the Empire, the Statistical Yearbooks of Brazil and the Demographic Censuses to construct new educational measures and calculate educational Gini indices between 1900 and 2000 for each region and for Brazil as a whole. Our results show that between

¹ Pesquisador – Insper Instituto de Ensino e Pesquisa – Endereço: Rua Quatá, 300 - Vila Olímpia/SP CEP: 04546-042 - E-mail: brunokk@uol.com.br - ORCID: <https://orcid.org/0000-0002-7731-3422>

² Professor – Insper Instituto de Ensino e Pesquisa e Faculdade de Economia, Administração e Contabilidade – Universidade de São Paulo – End.: Rua Quatá, 300 – Vila Olímpia – São Paulo/SP CEP: 04546-042 - E-mail: naercioamf@insper.edu.br - ORCID: <https://orcid.org/0000-0003-1241-6188>

³ Mestrando – Faculdade de Economia, Administração e Contabilidade – Universidade de São Paulo End.: Avenida Professor Luciano Gualberto, 908 – Butantã – São Paulo/SP. CEP: 05508-010 E-mail: pedro.ac@usp.br – ORCID: <https://orcid.org/0000-0001-5423-8856>

⁴ Analista de Dados, Development Impact Evaluation - DIME, The World Bank Group, Washington, DC, Estados Unidos – Endereço: 1818 H St NW, Washington, DC 20433 – E-mail: lvotti@worldbank.org ORCID: <https://orcid.org/0000-0001-5638-4566>

Recebido: 22/11/2017. Aceite: 06/06/2019.



Esta obra está licenciada com uma Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

1900 and 1930 the proportion of people with complete primary education in the population remained around 5%, while the share of secondary education was always below 1% and that only 0.3% had higher education complete. Thus, educational inequality remained constant until 1920, declining slowly between 1920 and 1950 and more rapidly thereafter.

Keywords

Education. Inequality. Gini Index.

JEL Classification

I24. I25.

1. Introdução

Apesar das importantes transformações sociais e no mercado de trabalho ocorridas no Brasil no período recente, que levaram à acentuada redução da desigualdade de renda na última década (Lustig, Lopez-Calva e Ortiz-Juarez 2013), o país figura ainda hoje entre os países com maior nível de desigualdade no mundo (Tsounta e Osueke 2014; CIA 2013). Além disso, os elevados níveis históricos de desigualdade social do Brasil são compartilhados entre alguns países latino americanos.

Um importante mecanismo de perpetuação da desigualdade de renda entre países se constitui nas diferenças de investimento e distribuição de capital humano (De Ferrenti et al. 2004). Nos países da América Latina, a educação formal se desenvolveu tardiamente, priorizando o acesso das elites, o que gerou uma escassez de mão de obra qualificada e desigualdades regionais (Camps e Engerman 2014). Após entrar na agenda de agentes da independência no século XIX, a expansão da educação à população foi esquecida com a consolidação de regimes conservadores no período pós-independência e permaneceu como privilégio dos mais ricos até o último quarto daquele século (Frankema 2009). Adicionalmente, as ondas de migrantes europeus ao Brasil entre o final do século XIX e início do século XX representaram choques positivos nos estoques de capital humano e reforçaram as desigualdades regionais de renda (Rocha, Ferraz e Soares 2015).

Este artigo pretende contribuir para a literatura com novas medidas de estoque educacional e também com uma análise detalhada da evolução da desigualdade educacional do Brasil no período que vai do início do século XX até a década de 1960. Nossa principal contribuição é a construção de medidas mais precisas de estoques da população por educação para a primeira metade do século XX, utilizando dados históricos dos Anuários Estatísticos do Brasil (AEBs) e dos Censos Demográficos. Além disso, empregamos um esforço para documentar as tendências e a dinâmica da desigualdade educacional nas regiões brasileiras ao longo de todo o século XX. Construímos o estoque inicial de indivíduos por grau de escolaridade para o início do século XX, a partir de dados pouco utilizados anteriormente, obtidos nos Relatórios Anexos ao Ministério dos Negócios do Império, o que nos permitiu estimar o índice de Gini educacional para o país como um todo e também para cada uma de suas regiões. A trajetória dessa medida de desigualdade educacional nos dá um panorama mais claro de como esta respondeu às reformas empreendidas pelo governo na tentativa de expandir o sistema de ensino. Além disso, o índice de Gini educacional nos permite avaliar como a desigualdade na educação se comportou em algumas regiões do país e em que medida a educação é responsável pela desigualdade regional existente ainda hoje.

A questão da desigualdade de renda vem ganhando destaque nas mais diversas áreas de pesquisa nos últimos anos, tendo em vista que estudos recentes apontam um aumento da desigualdade em todo o mundo (Dabla-Norris 2015). Há na literatura um conhecido debate sobre as possíveis causas para a alarmante desigualdade na América Latina. De um lado, a hipótese central é de que a atual desigualdade nos países latino-americanos tem raízes históricas e uma forte associação com a qualidade das instituições fundadas no período colonial e aquelas existentes hoje em dia (Acemoglu, Johnson e Robinson 2001; De Ferrenti et al. 2004). Nesse sentido, Engerman e Sokoloff (1994) também argumentam que as condições iniciais a partir das quais as colônias foram estabelecidas (incluindo a adequação do solo para culturas com grandes economias, produção com escravos e grande quantidade população nativa) tiveram impacto relevante sobre a evolução educacional e sobre a desigualdade social posterior. Do outro lado, outros autores argumentam que a alta desigualdade da América Latina é um fenômeno relativamente recente, tendo surgido em meio a transformações nos séculos XIX e XX, como o aumento dos preços internacionais das *commodities* e o isolacionismo dos países latino americanos após a Primeira Guerra Mundial (Williamson 2015). A evolução diferen-

ciada dos preços por tipo de *commodity* (em particular da cana de açúcar e do café) também parecer ser importante para explicar diferenças de desenvolvimento das regiões no Brasil (Leff 1972).

Além dessas correntes explicativas, a educação desempenha um papel relevante na perpetuação da desigualdade de renda. Os efeitos da desigualdade de capital humano sobre a desigualdade de renda foram bastante estudados na literatura (veja, por exemplo, Lam e Levison (1991), Lorel (2008), Barros (2011)). A transmissão intergeracional de escolaridade constitui o canal principal pelo qual a distribuição desigual de capital humano afeta a desigualdade de renda (Lorel 2008). A desigualdade educacional pode afetar a desigualdade de renda por meio da dispersão da escolaridade e dos efeitos da escolaridade sobre os rendimentos.¹ Adicionalmente, a educação pode afetar a desigualdade de renda através de externalidades positivas do capital humano, relacionadas à tecnologia (Barros 2011). Além do ganho de eficiência e aumento da produtividade dos indivíduos com maior nível de instrução, a tecnologia é desenvolvida respeitando as restrições dos outros fatores de produção. Portanto, regiões com maior disponibilidade de capital humano terão tecnologias que usam mais intensivamente este fator, e em consequência serão mais produtivas. Além disso, a importância da educação para a desigualdade é maior nas regiões menos desenvolvidas do Brasil, a saber a região Nordeste, uma vez que o diferencial salarial por grau de escolaridade nessas regiões é maior (Reis e Barros 1991).

Nesse sentido, a má distribuição de fatores de produção, especialmente do capital humano, acaba por impedir a convergência do nível de renda *per capita* entre as regiões, esperada por um modelo neoclássico (Barro e Sala-i-Martin 1992). Para o Brasil, cerca de 40% da desigualdade de renda observada nas décadas de 1980 e 1990 pode ser explicado pela variável educação (Ferreira 2000). Adicionalmente, a distribuição desigual de capital humano tende a persistir, uma vez que a acumulação de capital humano dos pais é positivamente relacionada à dos filhos. A ausência de políticas públicas que busquem redistribuir educação de forma mais igualitária entre as regiões, a desigualdade regional tende a se manter por tempo indeterminado (Barros 2011).

¹ Embora intuitivamente a desigualdade de escolaridade seja positivamente correlacionada com a desigualdade de renda, Lam, Leibbrandt e Finn (2015) argumentam que não há fundamento formal para que essa relação tenha esse sinal.

Nosso trabalho também dialoga com estudos que tentam explicar as causas da distribuição desigual de capital humano entre as regiões brasileiras, examinando a questão das diferenças de acesso à educação. A estrutura produtiva e de poder político das regiões parecem explicar parte da concentração da educação. O acesso à educação era restrito às elites locais até o final do século XIX. A expansão do acesso com as reformas educacionais no final do século XIX e início do século XX, além disso, beneficiou especialmente brancos (incluindo imigrantes europeus) e relegou nativos americanos, negros e mestiços ao segundo plano (Camps e Engerman 2014). Além disso, há evidências de que a interação entre a qualidade das instituições locais e o aumento dos preços internacionais das commodities afeta os níveis regionais de acumulação de capital humano (Musacchio, Martinez e Viarengo 2014).

Alguns estudos anteriores estimaram medidas de desigualdade educacional no longo prazo. Thomas, Wang e Fan (2001) calculam um índice de Gini educacional para diversos países a partir de 1960, e concluem que a tendência geral na segunda metade do século XX é de redução da desigualdade educacional, que é negativamente associada à média de anos de estudo e à riqueza do país. A associação do nível de escolaridade com a dispersão (medida pelo desvio padrão) segue o padrão de uma curva de Kuznets. Resultados semelhantes são encontrados por Lorel (2008) para o Brasil, entre 1950 e 2000, quando compara os estados brasileiros. Lee e Lee (2016) constroem uma medida de estoque educacional para o período entre 1870 e 2010 para 111 países, a partir de estimativas de taxas de matrícula no período entre 1820 e 1945 e das estimativas educacionais de Barro e Lee (2001) para o pós II Guerra. Van Leeuwen e Van Leeuwen-Li (2015) também calculam um índice de Gini educacional para diversos países, entre 1850 e 2000, incluindo o Brasil. Uma diferença importante entre nosso estudo e os de Lee e Lee (2016) e de Van Leeuwen e Van Leeuwen-Li (2015) está no grau de precisão dos dados utilizados, uma vez que as estimativas apresentadas por estes são mais gerais e para todo o país, enquanto as estimativas apresentadas neste trabalho são desagregadas por região e grau de escolaridade.

Nossa principal contribuição é estimar a evolução da desigualdade educacional no Brasil para a primeira metade do século XX, para quando não há microdados de pesquisas censitárias disponíveis. Nós adicionamos à literatura resultados obtidos a partir de fontes históricas não exploradas em outros estudos, como os relatórios estatísticos da antiga Diretoria Geral de

Estatísticas (DGE), relatórios estatísticos do final do século XIX e taxas de mortalidade calculadas separadamente por nível de escolaridade.

O artigo está dividido em seis partes, incluindo esta breve introdução. Na segunda parte, faremos uma descrição dos dados utilizados, bem como de suas fontes suas especificidades. Em seguida, explicaremos detalhadamente a metodologia utilizada no trabalho. Na quarta seção, realizamos um breve histórico da educação brasileira. Por fim, apresentamos os resultados obtidos e algumas conclusões sobre o assunto.

2. Dados Metodologia

Para estimar as estatísticas apresentadas neste artigo, foram combinadas diversas fontes históricas. Para estatísticas como taxa de alfabetização ou grau de escolaridade da população, utilizamos os Censos Demográficos, divulgados aproximadamente a cada dez anos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). O primeiro recenseamento nacional ocorreu em 1872, abrangendo todas as regiões do território, exceto o atual estado do Acre.

Também utilizamos os Anuários Estatísticos do Brasil (AEB) publicados pelo IBGE, na época chamado Diretoria Geral de Estatística (DGE) (Ministério da Agricultura, Indústria e Comércio, 1916; DGE, 1912-1936; Instituto Nacional de Estatística, 1936-1937; IBGE, 1938-1962). O primeiro AEB foi divulgado em 1916, contendo dados para os anos entre 1907-1912. Apenas a partir de 1936 a publicação passou a ser anual. Os Anuários reúnem uma grande quantidade de informações sobre educação, saúde, gasto público, trabalho, política, entre outros temas, entretanto, esses dados não são sempre comparáveis. Utilizamos as informações disponíveis acerca de matrículas, conclusões de curso e números de instituições de ensino.

Dados sobre taxa de alfabetização ou grau de escolaridade da população estão presentes nos Censos Demográficos em toda a série. Para estimar o percentual de indivíduos matriculados no nível escolar correspondente a cada faixa etária, cruzamos os dados dos AEBs com as estimativas

populacionais dos Censos Demográficos. Porém, devido às faixas etárias dos censos não serem exatamente compatíveis às de cada nível escolar, foi necessário realizar uma aproximação. Assim, para o ensino primário foi considerada a população de cada estado, entre 5 e 14 anos. Para o ensino secundário e técnico, o total de pessoas por estado entre 10 e 19 anos, e para o ensino superior indivíduos entre 20 e 24 anos.

Para 1930 utilizamos estimativas populacionais feitas pelo DGE, uma vez que nesta década não foi realizado Censo Demográfico. Entretanto, estas estimativas eram apenas para o total da população de cada estado. Para chegar à população por faixa etária, foi feita uma média entre o percentual da população em cada faixa etária por estado dos Censos de 1920 e 1940. Em seguida, este resultado foi multiplicado pela população total do estado correspondente em 1930.

Dados de matrículas no ensino primário, secundário e técnico no final do século XIX foram obtidos dos Relatórios Anexos ao Ministério dos Negócios do Império, da Diretoria Geral de Estatística (Directoria Geral de Estatística, 1875; Secretaria d'Estado dos Negocios no Imperio, 1887). Apesar de pouco conhecidos, estes relatórios apresentam uma vasta gama de informações sobre a educação no país naquele período, tais como receita e despesa com educação em nível provincial, número de estabelecimentos de ensino primário e secundário, públicos e particulares, frequência escolar em ambos os níveis e despesa por aluno.

Utilizamos dados de mortalidade da publicação Estatísticas do Século XX do IBGE para o Brasil, sem distinção de regiões. A única exceção é o estado de São Paulo, em que há informações mais precisas disponíveis no projeto Memória das Estatísticas Demográficas (SEADE 2010).

A partir dessas informações foi possível calcular um Gini educacional para o Brasil como um todo e para algumas de suas regiões. A estimação do Gini foi feita com base em estimativas da escolaridade da população brasileira ao longo do século XX. O estoque inicial, para o ano de 1900, foi calculado com base nos dados de matrículas e conclusões dos AEBs e de outros trabalhos e relatórios do DGE, especificados acima. Contamos também com dados de estoque populacional por escolaridade nos Censos Demográficos de 1940 a 1960.

2.1. Anuários Estatísticos

Para o cálculo dos estoques de nível de escolaridade utilizamos as informações acerca de matrículas, conclusões de curso e números de instituições de ensino contidas nos Anuários Estatísticos do Brasil (AEB) entre os anos de 1908 e 1962. Os anuários reúnem uma grande quantidade de informações sobre diferentes temas, educação, saúde, gasto público, trabalho, política, religião dentre outros. Apesar da riqueza de informações, os anuários tiveram publicações inconstantes nos anos iniciais e a disponibilidade de informações varia significativamente entre os anos e determinadas informações não são comparáveis entre diferentes anos. Para uma discussão detalhada da disponibilidade de informações nos AEBs ver Souto (2012), Guizzardi Filho et al. (2003) e IBGE (2006).

O primeiro AEB foi divulgado em 1916, contendo dados para os anos entre 1907-1912, mas a publicação foi interrompida entre 1913 e 1935. A partir de 1936, a publicação passa a ser anual, contudo, há inconstância nos dados sobre matrículas e conclusões. As Tabelas 1 e 2 abaixo sintetizam a disponibilidade de informações:

Tabela 1 – Dados disponíveis Anuários Estatísticos do Brasil - Matrículas

Ensino	Matrículas			
	Primário	Secundário	Superior	Profissionalizante
Série				
1907-1912	Não disponível	Disponível para todas as UFs exceto AC	Disponível para algumas UFs ou disponibilidade varia na série.	Disponível para todas as UFs exceto AC
1932-1937	Disponível para todas as UFs	Disponível para todas as UFs	Disponível para todas as UFs exceto AC	Disponível para todas as UFs
1938-1950	Não disponível	Não disponível	Não disponível	Não disponível
1951-1962	Não disponível	Não disponível	Não disponível	Não disponível

Fonte: Elaboração própria.

Tabela 2 – Dados disponíveis Anuários Estatísticos do Brasil - Concluintes

Ensino	Concluintes			
	Primário	Secundário	Superior	Profissionalizante
Série				
1907-1912	Não disponível	Disponível para algumas UFs ou disponibilidade varia na série.	Disponível para algumas UFs ou disponibilidade varia na série.	Disponível para todas as UFs exceto AC
1932-1937	Disponível para todas as UFs	Disponível para todas as UFs exceto AC	Disponível para todas as UFs exceto AC	Disponível para todas as UFs
1938-1950	Não disponível	Não disponível	Não disponível	Não disponível
1951-1962	Disponível para algumas UFs ou disponibilidade varia na série.	Disponível para algumas UFs ou disponibilidade varia na série.	Disponível para algumas UFs ou disponibilidade varia na série.	Disponível para algumas UFs ou disponibilidade varia na série.

Fonte: Elaboração própria.

Devido às interrupções na série, utilizamos métodos de interpolação linear para preencher lacunas, que serão explicados na próxima seção.

3. Metodologia

A metodologia deste artigo consiste em calcular o índice de Gini agregado com estimativas de estoques populacionais por nível de escolaridade ao longo do tempo. A principal dificuldade enfrentada é de que não há dados estatísticos ou censitários sobre o estoque da população por escolaridade no Brasil ou estado até o Censo Demográfico de 1940. Até então, a informação sobre a escolaridade disponível era somente de quantos sabiam ou não ler na população. Por esse motivo, para construir os estoques por escolaridade, partimos de estimativas do estoque de pessoas com cada nível de escolaridade em 1900 e, usando informações sobre os concluintes de cada nível, construímos a série para a primeira metade do século XX. Também, utilizamos a mesma metodologia de cálculo do Gini para os dados disponíveis nos censos demográficos de 1940 a 2000, com o intuito de

dar continuidade à série. Fazemos adicionalmente estimações separadas para as Unidades da Federação, o que nos permite agregar em regiões.

3.1. Metodologia das Estimativas da Distribuição da População por Escolaridade em 1900

Para a estimação do estoque inicial da população por escolaridade em 1900, nós nos inspiramos na metodologia de inventário perpétuo descrita por Földvári e Van Leeuwen (2014) para chegar a valores aproximados da participação relativa dos grupos de escolaridade na população de 5 anos ou mais em 1900. Nós trabalhamos com cinco categorias de escolaridade: sem escolaridade, ensino primário incompleto, ensino primário completo, ensino secundário completo (incluindo o ensino profissional), e ensino superior completo.²

Em linhas gerais, a metodologia consiste em obter a distribuição da população de uma determinada faixa etária por grau de ensino em algum período do século XIX, e aplicar esses percentuais para a mesma coorte em 1900. Por exemplo, se em 1880 a proporção de pessoas entre 5 e 10 anos de idade matriculadas no primário fosse de 10%, então supomos que o mesmo percentual de pessoas entre 25 e 30 anos de idade em 1900 certamente possua o ensino primário. Nas estimativas populacionais do século XIX, consideramos também a população escrava nos cálculos (sem nenhuma escolaridade).

Há duas hipóteses implícitas para a realização dessa estimativa. Em primeiro lugar, supomos que todos os matriculados em um grau escolar de uma coorte irão se formar (taxa de evasão nula). Essa hipótese será relaxada mais adiante, porém os resultados não se alteraram de forma relevante. A segunda hipótese é de que a taxa de mortalidade e migração daqueles é homogênea entre pessoas com diferentes graus escolares. Apesar de forte, não é possível estimar taxas de mortalidade e migração diferenciadas por grau escolar com os dados disponíveis. Somente estudos utilizando dados muito recentes são capazes de realizar esse tipo de estimação de forma

² O ensino primário equivale ao que atualmente é chamado de primeiro ciclo do ensino fundamental. O secundário constitui um grau de ensino que compreende o segundo ciclo do ensino fundamental e o ensino médio. Para essa análise, consideramos o ensino profissional como parte do ensino médio.

precisa, porém, fizemos um teste de robustez com estimativas de mortalidade diferenciadas entre pessoas alfabetizadas e analfabetas.³

Para obter as proporções de matriculados por faixas etárias, calculamos as proporções de crianças e jovens matriculados no primário e no secundário em relação ao total de indivíduos nas faixas de idade consideradas adequadas para cada grau. Essa estimação foi possível para o início dos anos 1870 (com dados do Recenseamento de 1872 e dos Relatórios e Trabalhos Estatísticos de 1875) e dos anos 1880 (com dados dos Censos de 1872 e 1890, e dos Trabalhos da Seção de Estatística de 1886). Para a década de 1880, também foi possível estimar a proporção de adultos matriculados no ensino superior.⁴

Com as estimativas de taxas de matrículas em mãos, aplicamos essas proporções de pessoas com primário completo, secundário completo e ensino superior completo para a população de 1900 por estado da seguinte forma. As proporções de pessoas com primário completo em 1880 foram aplicadas para as faixas com 10 a 29 anos em 1900, enquanto para as coortes mais velhas no mesmo ano, aplicamos as proporções de 1870. As proporções de pessoas com secundário completo de 1880 foram aplicadas às pessoas entre 15 e 29 anos em 1900, e para os mais velhos aplicamos as proporções de 1870. Por fim, as proporções de pessoas com ensino superior foram aplicadas à população com 25 anos ou mais em 1900. Como aqueles que fizeram um grau escolar mais elevado passaram obrigatoriamente pelos graus mais elementares, evitamos que eles fossem contados duplamente nos graus escolares anteriores.

Entre aqueles que não foram designados a algum grau escolar, dividimos entre aqueles que sabiam ou não ler. O número de analfabetos foi obtido a partir do Censo Demográfico de 1900, e esse grupo foi considerado sem escolaridade. Os demais, que sabiam ler, porém não se encaixaram em nenhum dos graus escolares completos, foram considerados como tendo primário incompleto.

³ Há alguma evidência de que as diferenças de mortalidade entre os mais e menos educados não era tão grande no início do século XX. Cutler, Deaton e Lleras-Muney (2006) mostram que o gradiente de mortalidade por renda/educação aumenta com o tempo, com a introdução de conhecimento relevante à saúde, porque a incorporação desse conhecimento em práticas que de fato aumentam a expectativa de vida é mais rápida entre os mais escolarizados. Por exemplo, o gradiente era baixo entre aristocratas e pessoas comuns na Inglaterra antes Iluminismo, ou entre filhos de pais médicos ou não médicos antes da descoberta de que algumas doenças eram causadas pela ação de germes.

⁴ Detalhamos os procedimentos para obtenção dos percentuais de adultos com ensino superior no Apêndice.

Como mencionamos anteriormente, é possível relaxar a hipótese de que todos os matriculados se formam. Se tivermos a proporção dos concluintes em relação aos matriculados, podemos obter uma estimativa mais real do número de formados em um grau escolar. Dessa forma, tentamos lidar com o problema aplicando a taxa de conclusão aos matriculados e multiplicando o valor resultante (número de concluintes em um ano) pelo tamanho da faixa etária que frequentaria aquele grau escolar.

Os primeiros dados de concluintes para todos os níveis de ensino estão disponíveis somente a partir da década de 1930. No entanto, como a taxa de conclusão por matriculado é relativamente estável entre 1932 e 1949, utilizamos essa taxa para estimar a proporção de concluintes por faixa etária em 1870 e 1880. Em seguida, calculamos novamente as proporções de formados em 1900, como descrito anteriormente. Apesar de cada alteração implementada para lidar com um problema diferente gerar modificações na distribuição de pessoas por escolaridade, essas distribuições não se alteram significativamente.

3.2. Composição Educacional e População ao Longo do Século XX

Para estimar os estoques populacionais por escolaridade, tratamos de forma diferenciada os grupos com níveis de ensino completos (primário completo, secundário completo e superior completo) de pessoas sem escolaridade ou com primário incompleto. Para estimar o número de pessoas com cada nível de ensino completo, partimos da estimativa do estoque inicial de 1900 e, iterativamente para cada ano subsequente, descontamos os óbitos e somamos o número de concluintes, subtraindo o número de concluintes do nível seguinte. Por exemplo, o estoque total de pessoas com ensino primário em um determinado ano é calculado como a soma entre o estoque do ano anterior e o número de concluintes do ensino primário daquele ano menos os concluintes do ensino secundário do mesmo ano.

Para um grau escolar j no ano t , o estoque populacional será dado por:

$$\text{Pop}_{jt} = \text{Pop}_{j,t-1} \times (1 - \text{Taxa de Mortalidade}_{jt}) + (\text{Concl}_{jt} - \text{Concl}_{j+1,t}) \quad (1)$$

onde Pop_{jt} é a população com escolaridade j no ano t , Taxa de Mortalidade $_{jt}$ é a taxa de mortalidade e $Concl_{jt}$ é o número de concluintes no grau j no ano t .

A disponibilidade de dados de concluintes varia bastante ao longo dos anos, mas estão disponíveis para quase todas as Unidades da Federação entre os anos de 1907 a 1912, 1932 a 1949 e 1951 a 1962. As séries de concluintes para cada nível educacional foram interpoladas para os anos em que os AEBs não foram publicados, ou em que estes não apresentam essa informação. Assumiu-se que a quantidade de matrículas e as taxas de conclusão pouco mudam entre os anos de 1899 e 1908, para os estados em que existem informações sobre o número de concluintes em 1908. Para os estados em que não há informações do número de concluintes em 1908, fizemos uma interpolação linear ponto a ponto, o que evita a existência de quinas, supondo que o número de concluintes de cada nível de ensino era zero em 1899. Dessa forma, é possível construir uma série de concluintes de cada nível educacional ininterrupta entre 1900 e 1960. Para que as estimativas fossem comparáveis com os censos de 1940 e 1950, os concluintes dos níveis secundário e profissional foram agrupados para estimar o estoque de pessoas com ensino médio.

Para as pessoas sem escolaridade e o primário incompleto, procedemos da seguinte forma. O estoque da população sem escolaridade é dado imediatamente pelo número de pessoas que não sabem ler. Essa informação e o dado da população total estão disponíveis em todos os Censos Demográficos, desde 1872, para o Brasil como um todo e para as Unidades Federativas. Para preencher a população total e as subpopulações nos anos intercensitários, foi utilizada uma interpolação linear simples da proporção de indivíduos. Dessa forma, faltaria apenas estimar o estoque da população com ensino primário incompleto. Calculamos esse estoque como a diferença entre a população total e os demais estoques já estimados.

Adicionalmente, os óbitos foram calculados a partir das informações sobre a taxa de mortalidade. A taxa de mortalidade média do Brasil foi obtida dos Censos Demográficos (IBGE 2007) e assumiu-se que a taxa é a mesma entre todas as Unidades da Federação, com exceção de São Paulo, cuja taxa de mortalidade para o período analisado é divulgada pela Fundação SEADE (SEADE 2010).

É razoável pensar que a taxa de mortalidade varia de acordo com o nível de instrução (que é positivamente correlacionado com a renda). Para tentar obter uma taxa de mortalidade diferenciada por instrução, utilizamos os dados de alfabetização dos indivíduos por faixa etária, disponíveis nos Censos Demográficos de 1940 e 1950, para calcular a mortalidade relativa. Consideramos no cálculo indivíduos com 40 anos ou mais, uma vez que até esta faixa etária ainda havia uma parcela considerável da população se alfabetizando. Para o cálculo, agregamos separadamente indivíduos que sabiam ou não ler, com entre 40 e 69 anos no Censo de 1940 e comparamos com os grupos correspondentes de indivíduos com entre 50 e 79 anos no Censo de 1950. A taxa é calculada dividindo o número de pessoas por faixa etária no período final pelo número de pessoas na faixa etária anterior no período inicial. Com isso, obtivemos os resultados apresentados na Tabela 3.

Tabela 3 – Taxas de mortalidade decenais

	Comparação			Agregado
	40 a 19 / 50 a 59	50 a 59 / 60 a 69	60 a 69 / 70 a 79	
Alfabetizados	-21.41%	-30.01%	-49.72%	-28.25%
Analfabetos	-24.07%	-28.32%	-49.10%	-29.78%
Total	-22.96%	-29.00%	-49.33%	-29.16%

Fonte: Censos Demográficos/IBGE. Elaboração Própria. Ainda, um ajuste foi feito no período inicial da comparação, incluindo nos alfabetizados e subtraindo dos analfabetos, os indivíduos que na data da pesquisa estavam frequentando a escola, porém ainda eram analfabetos.

O diferencial a partir da taxa média de -29,16% é de 0.97 para os alfabetizados e de 1.02 para os analfabetos.

3.3. Gini Educacional

Para cada categoria educacional das cinco utilizadas, atribuiu-se a quantidade de anos de educação formal equivalente à duração de todos os níveis concluídos pelos indivíduos presentes no grupo. Assim, para o ensino primário consideramos 4 anos de estudo; para o ensino médio, 11 anos de estudo; e 15 anos de estudo para o ensino superior. Dessa forma, como os anos de educação são constantes para cada grupo, não é possível mensurar a dispersão de escolaridade intragrupo. Para os indivíduos analfabetos

optamos por atribuir uma quantidade de anos de estudo igual a zero (0 anos). Também, para aqueles com o nível primário incompleto, devido à falta de informações no período, atribuiu-se um valor *ad hoc* de 1 ano de estudo formal.

Por ser uma medida comumente utilizada para medir desigualdade e de fácil interpretação, optou-se pelo índice de Gini para medir a desigualdade na distribuição de educação. A metodologia usada é uma adaptação da forma usual de cálculo da medida (Cowell 2009). Como as informações existentes sobre o nível de escolaridade para o período analisado e nossas estimações de estoque de pessoas por nível de ensino estão agregadas, isto é, não temos informações em nível individual, adaptamos a fórmula do índice de Gini para a estrutura de nossos dados. A forma de cálculo é muito similar à utilizada por Thomas, Wang e Fan (2001).

O Gini educacional (G) para um determinado ano e região é calculado como a média da diferença entre todos os possíveis pares de indivíduos, dividida pela média de anos de escolaridade (\bar{X}). Como há apenas cinco tipos distintos de indivíduos, a fórmula pode ser simplificada da seguinte maneira:

$$G = \frac{\sum_{i=1}^5 \sum_{j=1}^5 Q_i Q_j d_{ij}}{2N^2 \bar{X}} \quad (2)$$

Onde N é a população com 5 anos ou mais; \bar{X} é a média de anos de estudo; Q_i é a quantidade de pessoas do grupo i ; de maneira análoga Q_j é a quantidade de pessoas do grupo j ; e d_{ij} é a diferença de anos de estudo entre os grupos i e j .

4. Breve Histórico da Educação no Brasil

Durante boa parte do século XIX, o Brasil – assim como outros países latino-americanos – esteve entre os países com os piores indicadores educacionais do mundo, em termos de nível e desigualdade escolar (Lustig, Lopez-Calva e Ortiz-Juarez 2013).

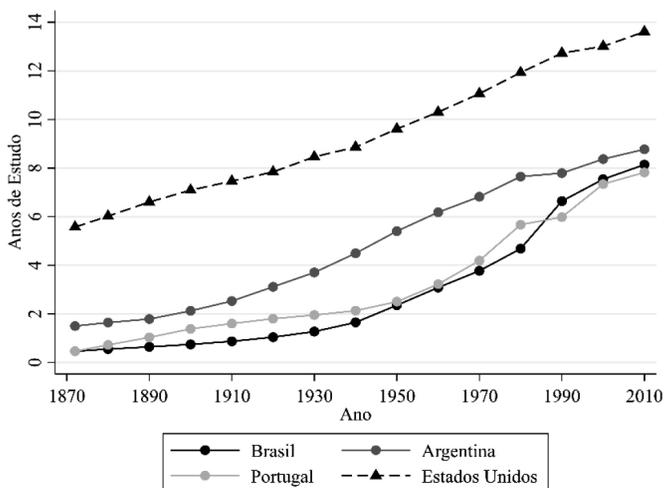


Figura 1 - Média de anos de estudo para países selecionados

Fonte: Clio-Infra (Leeuwen, Van Leeuwen-Li e Földvari 2013). Elaboração própria. Dados para população de 15 anos ou mais.

A Figura 1 compara a evolução da média de anos de estudo para quatro países: Brasil, Argentina, Estados Unidos e Portugal. Percebe-se que a média de anos de estudo no Brasil era muito baixa no final do século XIX (0,7 anos de estudo em 1900) e assim permaneceu durante as três primeiras décadas do século XX. A partir de 1940, este indicador passou a apresentar crescimento mais acelerado no Brasil, tornando a curva mais inclinada. Somente na década de 1990 é que a média de anos de estudo do Brasil chega ao nível observado nos Estados Unidos no início do século XX, o que explicita o grande atraso educacional do país. As trajetórias do indicador na Argentina foram bem semelhantes à do Brasil, porém num patamar mais elevado, com crescimento comparativamente mais equilibrado ao longo de todo o período, exceto por uma década de estagnação entre 1980 e 1990, década em que o Brasil mais avançou nesse indicador.

Há diversos fatores que explicam o atraso relativo do Brasil. De acordo com Colistete (2016), em contraste com outras colônias, como os Estados Unidos e Canadá, o padrão educacional observado nas colônias latino americanas seguia o das metrópoles, sendo extremamente restrito, focado no ensino secundário e superior e voltados para as elites.

Além disso, no Brasil havia um descaso político com o ensino primário, que se refletia no baixo nível de gastos por aluno (dez vezes inferior ao dos Estados Unidos).

Mesmo em comparação aos demais países da América Latina, o Brasil iniciou tardiamente uma expansão do acesso ao ensino básico, como fica evidenciado pela baixa média de anos de estudo na Figura 1. Apenas nas décadas finais do Império é que se iniciou um lento processo de difusão de escolas públicas e aumento do número de alunos matriculados (Marcílio 2005, 115).

Durante o período conhecido como Primeira República, ocorreram diversas mudanças no campo educacional. Em primeiro lugar, a Constituição Republicana de 1891 estabeleceu que apenas indivíduos alfabetizados tinham direito ao voto, portanto, somente uma parcela reduzida da população seria considerada apta a votar. Dessa forma, fornecer maior acesso à educação passou a ser também uma questão política, uma vez que não somente satisfaria as demandas da população, como tornaria possível angariar um maior número de eleitores.

Além disso, esta nova Constituição descentralizava a administração do sistema educacional, tornando-a responsabilidade dos estados e municípios. O fato de o governo brasileiro ter delegado a administração do ensino aos estados durante boa parte da história educacional é incluído como uma das causas para o atraso educacional do país (Colistete 2016) e para as desigualdades regionais, pois os estados não possuíam o mesmo volume de recursos e apenas o governo central conseguiria arrecadar os recursos necessários para democratizar a educação. Dessa forma, os estados mais desenvolvidos e urbanizados tiveram um progresso educacional sensivelmente mais rápido (como é o caso de São Paulo e outros estados das regiões Sul e Sudeste, como é possível notar nas Figuras 5 e 6).

Ainda, durante a Primeira República, o sistema educacional brasileiro pouco evoluiu e as políticas educacionais adotadas pelos governantes do período foram, em grande parte, controversas e pouco efetivas levando a um baixo crescimento nas taxas de matrículas e conclusões, crescimento este concentrado nas regiões mais ricas e urbanizadas, como pode ser visto nas Figuras 3 e 4. De acordo com Kang (2010), isso pode ser explicado pela distribuição altamente desigual de poder político no período. Ele argumenta que uma expansão da educação pública significativa dependeria de uma

maior participação política das classes que mais seriam beneficiadas por esta expansão, isto é, os mais pobres. Aumentando a participação política da população, através de suspensão das restrições ao voto, os indivíduos teriam maior poder para demandar educação pública, de forma que a alfabetização da população, bem como a taxa de matrículas, elevaria.

Após a crise 1929 e seus reflexos para a economia brasileira, o Estado adotou uma postura mais intervencionista, especialmente no que diz respeito à educação. Somando-se a isto à Revolução de 1930, uma nova reforma do ensino foi proposta, com a intenção de criar, pela primeira vez, um sistema nacional de ensino. Em 1931 foi criado o Ministério da Educação e Saúde, e logo se seguiu a criação da primeira universidade pública, em 1934, localizada em São Paulo. A partir daí, o número de universidades cresceu lenta e continuamente. No que se refere ao ensino secundário, a reforma estabeleceu diretrizes mais claras que nas décadas anteriores. O ensino profissional também se expandiu no período, estimulado pelo processo de substituição de importações, que se intensificaria futuramente.

O ensino secundário entrou em foco nesse período, o que é evidenciado no painel b) da Figura 3, onde a taxa de matrícula neste nível de ensino, que se mantivera praticamente a mesma entre 1907 e 1932, passa a crescer a partir de meados da década de 1930. Todavia, ainda neste momento a pequena expansão do ensino secundário era voltada principalmente às elites, ao passo que para as classes sociais menos favorecidas surgia como opção o ensino médio profissionalizante, representado principalmente pelo ensino industrial, o que promovia um processo de segmentação social da educação (Marcílio 2005).

Na década de 1940, o ministério da educação volta a implementar reformas parciais, chamadas de Leis Orgânicas. Essas reformas tangenciavam principalmente o ensino primário e secundário e, em alguns casos, cursos do ensino profissional, como é o caso das Leis Orgânicas de 1942 e 1946. Estas passaram a reconhecer os cursos Industrial, Comercial, Normal e Agrícola como sendo de nível secundário. Seus efeitos podem ser notados na Figura 6, onde se registra um grande aumento no número de indivíduos com ensino primário e médio entre 1940 e 1950. Ainda, influenciado por propostas internas e externas, no período entre 1945 e 1964, o sistema educacional brasileiro passou por importantes transformações. Foram criadas campanhas com o intuito de promover a alfabetização de adultos, e a expansão do ensino primário e superior.

Lee e Lee (2016) pontuam que o período de crescimento mais acelerado das matrículas, em nível mundial, deu-se somente após a Segunda Guerra Mundial. No período anterior, apenas os países mais desenvolvidos, que implementaram políticas de escolarização básica obrigatória entre os séculos XVIII e XIX, viram uma expansão real do número de indivíduos matriculados, em especial nos ensinos primário e secundário. No que tange aos países menos desenvolvidos, o crescimento da taxa de matrículas ganhou relevância somente após a década de 1950, o que também pode ser visto na Figura 1, onde o crescimento da média de anos de estudo do Brasil fica mais evidente após a década de 1940.

Além do atraso relativo do Brasil, também foram constatadas disparidades regionais em educação internas ao país, contribuindo para geração de desigualdade social. A Figura 2 ilustra esse ponto usando a taxa de alfabetização da população. Apesar do grande aumento da proporção de alfabetizados em todo o país, entre o final do séc. XIX e o começo do século XXI, verifica-se um relativo atraso de algumas regiões. No Censo de 1872, as regiões com maior proporção de alfabetizados eram Norte e Sul, enquanto não havia grande distinção entre Sudeste, Nordeste e Centro-Oeste. Antes do final do século XIX, a região Sudeste já apresenta acelerado crescimento em comparação com as demais. A região Centro-Oeste, no entanto, só chega ao nível de alfabetização das regiões que em 1872 eram mais avançadas depois de 1900, enquanto a região Nordeste só chega a esse nível depois de 1920. Essa região permanece com a menor taxa de alfabetização no período e mesmo no ano final, 2010, a ela chega a um nível que foi atingido pelas regiões mais alfabetizadas já em 1980.

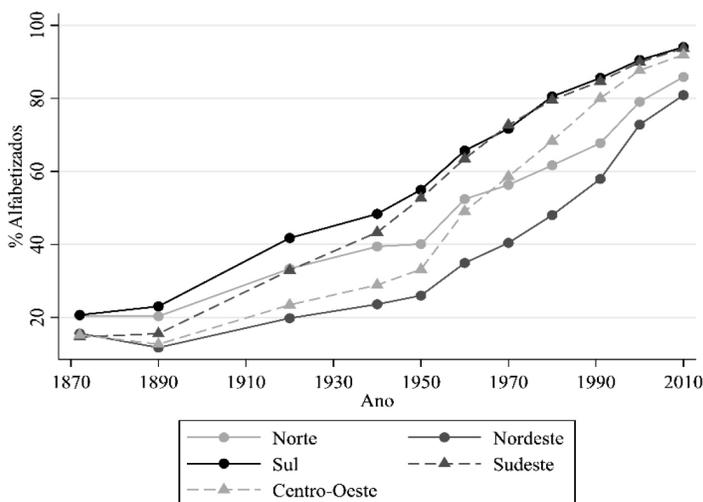


Figura 2 - Alfabetização nas Regiões Brasileiras

Fonte: Censos Demográficos/IBGE. Elaboração própria. Dados para a população de 5 anos ou mais. Obs.: o Censo de 1900 omitiu alguns distritos da pesquisa, e por esse motivo não o incluímos nesse gráfico.

Esse panorama geral da educação no Brasil reflete políticas educacionais ao longo de todo o século XX, que afetam as matrículas e conclusões nos graus escolares. Essas taxas dão uma visão do comportamento do sistema educacional brasileiro e expõem o baixíssimo nível de matrículas em todos os graus escolares e o ainda menor nível de conclusões. Também chama atenção quão pequena foi a evolução do acesso à educação no período, em especial no ensino superior, que em 1960 possuía menos de 2% da população matriculada.

A Figura 3 mostra a proporção de matriculados em cada grau escolar e a população na faixa etária considerada ideal para o nível. Todas as regiões brasileiras apresentam um crescimento das taxas de matrícula em todos os graus escolares, ocorrendo primeiro e de forma mais acentuada no ensino primário.⁵

⁵ A Lei 5.692/1971, conhecida como 2ª Lei de Diretrizes e Bases da Educação Nacional, promoveu mudanças na organização dos graus escolares da educação básica, substituindo o primário de quatro anos e o secundário de sete anos (quatro anos no 1º ciclo e três anos do 2º ciclo) pelo sistema com um primeiro grau de oito anos e um segundo grau de três anos (Kang 2018). Apesar dessas mudanças explicarem parte dos movimentos verificados nas Figuras 3 e 4 após 1970, elas não alteram nossas estimativas apresentadas nas Figuras 6 a 12. Como descrito na seção 3.2, nessas estimativas, nós

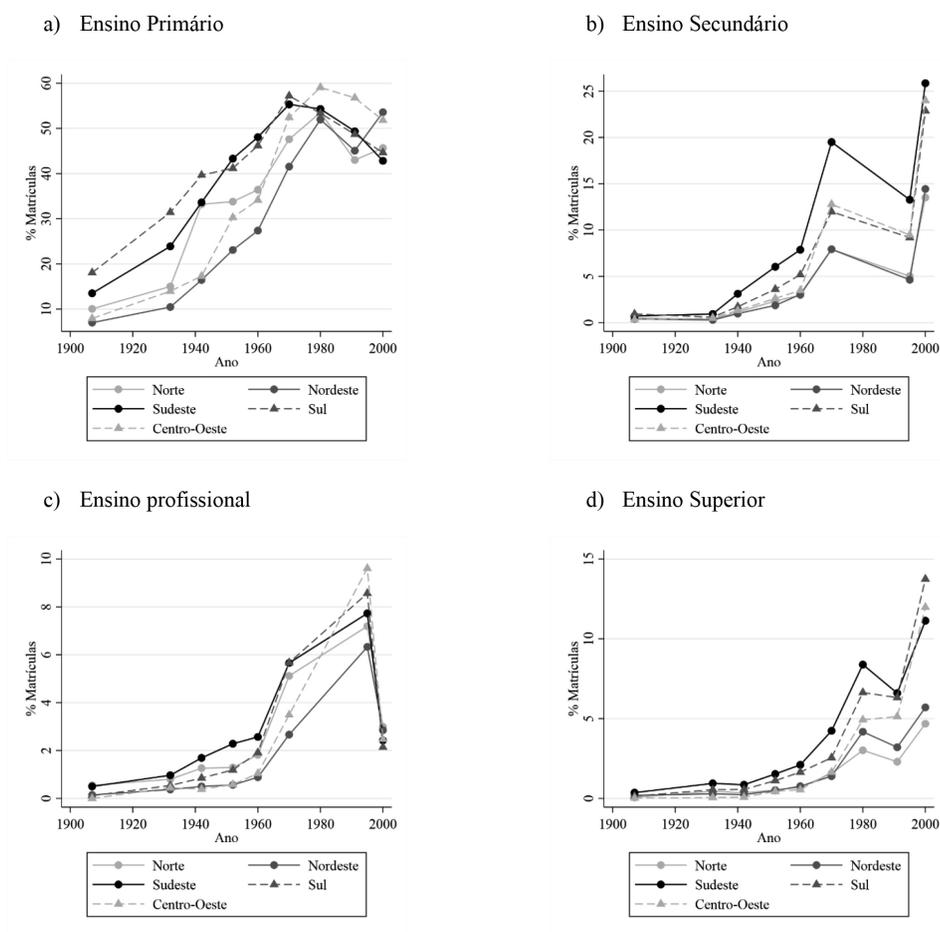


Figura 3 – Taxas de Matrícula

Fonte: Os dados de matrículas são provenientes dos Anuários Estatísticos do Brasil, de diferentes anos (DGE 1912-1936; IBGE 1932-1980) e do Censo Escolar de 1995 e 2000 para os anos posteriores à 1990. Para se chegar à taxa de matrícula, dividimos o número de indivíduos matriculados pelo total da população na faixa etária aproximada de frequência ideal de cada nível escolar (Primário: 5 a 14 anos; Secundário: 10 a 19 anos; Profissional: 10 a 19 anos; Superior: 20 a 24 anos).

O ensino secundário e o profissional possuem níveis iniciais comparativamente muito mais baixos e, em todas as regiões, o primeiro só começou a apresentar crescimento a partir de 1932. No caso do ensino superior,

consideramos agregados compatíveis com o sistema anterior àquela Lei.

apesar de aumento significativo nas primeiras décadas do século XX nas regiões Norte, Sul e Sudeste, só há uma tendência de crescimento sistemático da taxa nas demais regiões a partir de 1942.

Comparando as regiões, é perceptível que as regiões Sul e Sudeste apresentam as maiores taxas em quase todos os níveis de ensino. Particularmente, o crescimento da taxa para essas duas regiões no nível superior é expressivo a partir da década de 1940. Em contrapartida, a região nordeste apresenta as menores taxas em toda a série para o ensino primário e, com exceção de alguns pontos, também nos níveis secundário e profissional, mas com níveis de matrícula ainda muito baixos. As regiões Norte e Centro-Oeste têm taxas medianas em comparação com as demais regiões para os níveis primário e secundário, mas em particular, a região centro-oeste tem as piores taxas de matrícula no ensino superior durante o período.

Na Figura 4 apresentamos as taxas de conclusão, calculadas como a proporção de concluintes em relação à população na faixa de idade correspondente ao grau escolar. A taxa de conclusão reflete de forma mais precisa o incremento de um grau escolar em uma faixa de idade, porém, exceto pelo ensino superior, cuja série se inicia em 1912, os demais graus escolares possuem dados disponíveis somente a partir de 1932.

Enquanto as regiões Sul e Sudeste apresentam crescimento entre 1932 e 1942, a taxa nas demais regiões permanece praticamente constante. A região Norte é aquela cuja taxa mais destoia do comportamento da curva de taxa de matrículas, o que é explicado em parte pela imprecisão das informações, em particular para locais mais distantes do centro político e intelectual do país e com maiores dificuldades de acesso. Na região Centro-Oeste, a queda no percentual de concluintes entre as décadas de 1950 e 1960 pode ter sido causada por fatores demográficos, uma vez que na metade final da década de 1950, a cidade de Brasília foi construída, causando um grande aumento do número de habitantes da região. De maneira semelhante à taxa de matrículas, a taxa de conclusões para os ensinos secundário, profissional e superior são muito pequenas em todas as regiões brasileiras. O ensino superior também apresenta crescimento das taxas de conclusão das regiões Sul e Sudeste com relação às demais após 1940.⁶

⁶ Apresentamos no Apêndice gráficos com o número indivíduos na faixa etária correspondente a cada grau de ensino dividido pelo número de escolas deste mesmo grau.

Comparando a taxa de conclusão à taxa de matrícula é possível ter uma noção da magnitude da taxa de evasão escolar no período, principalmente no ensino primário. Isto se deve ao fato de que a principal barreira para a continuidade dos estudos se encontrava na transição entre o ensino primário e o ensino secundário, de forma que boa parte dos indivíduos que abandonavam os estudos o faziam ainda no ensino primário (Marcílio 2005).

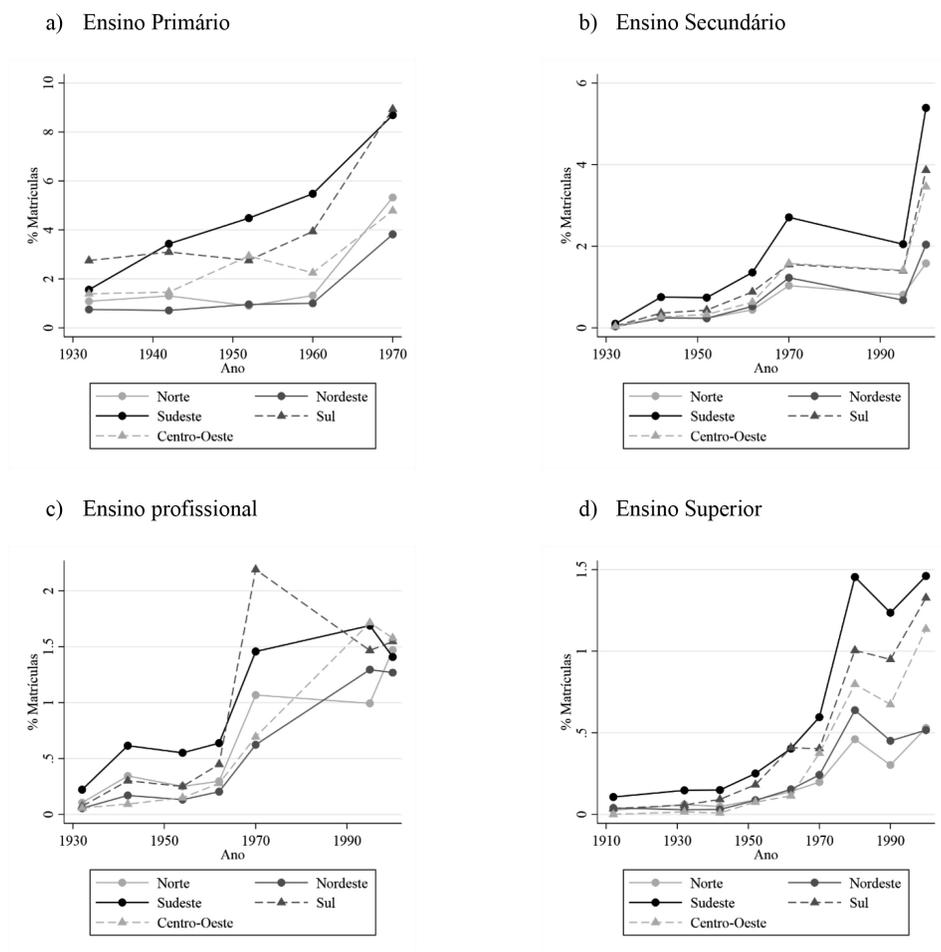


Figura 4 – Taxas de Conclusão

Fonte: Os dados de conclusão de curso são provenientes dos Anuários Estatísticos do Brasil de diferentes anos (DGE 1912-1936; IBGE 1932-1990) e do Censo Escolar de 1995 e 2000 para os anos posteriores à 1990. Para se chegar à taxa de conclusão, dividimos o número de concluintes pelo total da população na faixa etária aproximada de frequência ideal de cada nível escolar (Primário: 5 a 14; Secundário: 10 a 19; Profissional: 10 a 19; Superior: 20 a 24).

Os resultados acima descritos podem, em parte, ser explicados pela forma de organização e evolução do sistema educacional brasileiro. Inicialmente, antes do século XX, o sistema educacional brasileiro evoluiu muito lentamente, principalmente durante o período em que o país foi colônia portuguesa.

5. Resultados

Em primeiro lugar, a Figura 6 mostra nossa primeira contribuição empírica principal, os estoques populacionais por escolaridade estimados para o Brasil a partir dos níveis iniciais calculados para o ano de 1900. As linhas representam os níveis estimados de pessoas em cada grau escolar (primário, secundário e ensino superior), enquanto os pontos mostram os totais obtidos a partir dos Censos Demográficos de 1950 a 1960.

Para o ensino primário, maioria da população educada no período, nossa estimativa é semelhante ao observado no Censo Demográfico de 1950, mas em 1960 subestimamos o número de pessoas com esse nível de ensino. Uma possível explicação para isso é que no Censo Demográfico deste ano os indivíduos declararam sua escolaridade independente da modalidade de ensino que haviam cursado, ao passo que os dados utilizados para estimar a população com ensino primário neste período, provenientes dos AEBs, não incluíam aqueles que haviam cursado o ensino supletivo, restringindo-se apenas aos indivíduos que frequentavam o ensino primário fundamental comum.

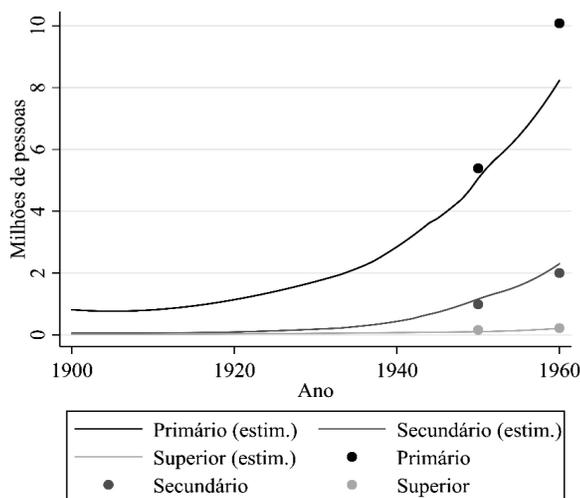


Figura 6 – Número de Pessoas com 5 Anos ou Mais por Grau Escolar

Fonte: Elaboração própria. As linhas representam as estimativas para todo o período e os pontos representam os dados observados nos Censos Demográficos (IBGE 1940-1960).

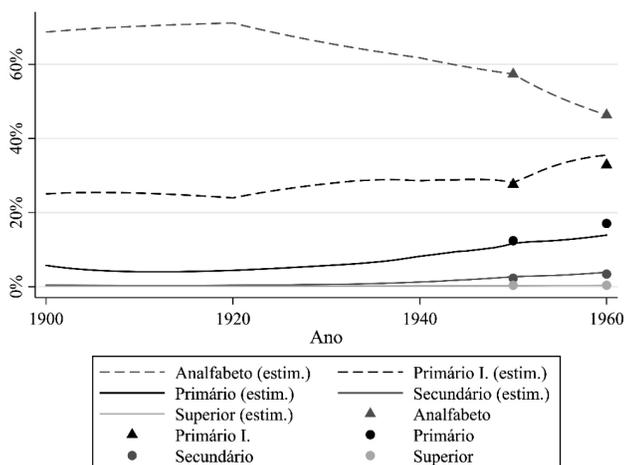


Figura 7 – Composição Educacional do Brasil

Fonte: Elaboração própria. As linhas representam as estimativas para todo o período e os pontos representam os dados observados nos Censos Demográficos (IBGE 1940-1960).

Para os outros dois níveis de ensino, nossas estimativas estão mais próximas aos níveis observados em 1950 e 1960, o que aumenta nossa confiança sobre elas para a primeira metade do século XX. Como mostra a Figura 6, a quantidade de pessoas com o primário completo é muito superior àque-la dos demais níveis educacionais, e apresenta crescimento substancial a partir de 1940. A Figura 7 mostra que a parcela da população com ensino primário, secundário e superior cresceram muito lentamente ao longo da primeira metade do século XX. A proporção de pessoas com ensino primário completo na população permaneceu em torno de 5% entre 1900 e 1930. Nesse período, a proporção de pessoas com ensino médio completo permaneceu sempre abaixo de 1%, e a proporção com ensino superior completo estava abaixo de 0.3%.

Nosso segundo resultado principal é a estimação do índice de Gini educacional durante a primeira metade do século XX, para o Brasil e quatro das cinco regiões brasileiras, que mostramos nas Figuras 8 e 9. A região Norte foi excluída da análise devido à baixa qualidade de informações para o começo do século XX, porém, no cálculo do índice para o Brasil estão incorporadas todas as cinco regiões. Adicionalmente, é utilizada a mesma metodologia de agregação em cinco grupos de escolaridade para os dados dos Censos de 1950 a 2000 para fins de comparação.

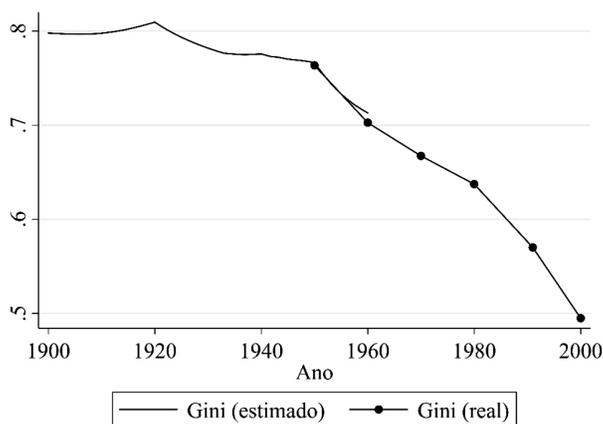


Figura 8 – Gini Educacional Estimado vs Real – Brasil

Fonte: Elaboração própria. As linhas para o período entre 1900 e 1960 representam as estimativas do Gini educacional feitas a partir da composição educacional anteriormente estimada. Os pontos entre 1950 e 2000 representam o Gini educacional calculado a partir dos dados dos Censos Demográficos (IBGE 1950-2000).

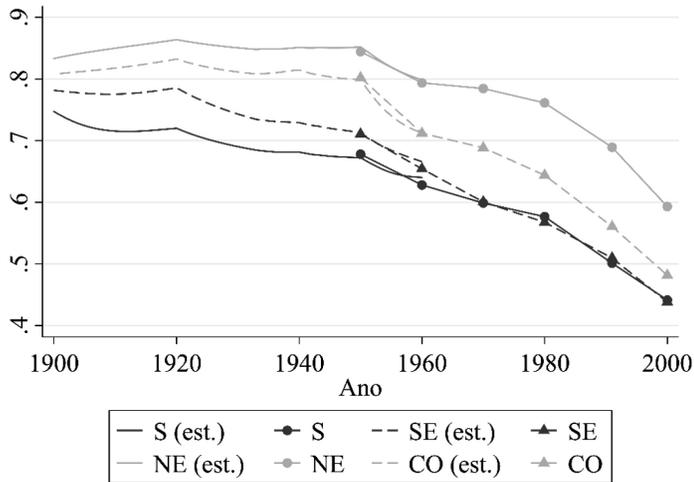


Figura 9 – Gini Educacional Estimado vs Real por Região

Fonte: Elaboração própria. As linhas para o período entre 1900 e 1960 representam as estimativas do Gini educacional feitas a partir da composição educacional anteriormente estimada. Os pontos entre 1950 e 2000 representam o Gini educacional calculado a partir dos dados dos Censos Demográficos (IBGE 1950-2000).

Em primeiro lugar, nossas estimativas chegam a 1950 e 1960 muito próximas aos níveis obtidos a partir dos estoques populacionais por educação observados nos dados dos Censos Demográficos, o que nos traz confiança sobre os movimentos do indicador na primeira metade do século XX.

Como esperado, o nível de desigualdade educacional em 1900 é bastante elevado e, ao invés de cair com o aumento de indivíduos escolarizados, a desigualdade educacional brasileira se mantém praticamente estável em um nível elevado durante boa parte da primeira metade do século XX. Na Figura 9, mostramos os índices separadamente por região. Em 1900, as quatro séries estão relativamente próximas entre si (entre 0,73 para a região Sul e 0,80 para as demais). A desigualdade interna às regiões tem variação perceptível ao longo do tempo. Nas primeiras décadas do século, apenas a região Sul apresenta declínio no índice de Gini, enquanto as outras três regiões analisadas e o Brasil apresentam um aumento de desigualdade. Notadamente, a região Nordeste mantém-se com a maior desigualdade durante todo o período analisado. A região Sul permanece até a o fim da década de 1950 com a menor desigualdade, mas a região

Sudeste, que tem queda expressiva desde a década de 1920, aproxima-se consideravelmente desde meados do século. A queda de desigualdade para as regiões Centro-Oeste e Nordeste, entretanto, passa a ser expressiva apenas no início da década de 1950. A região Centro-Oeste, que até a metade do século estava consideravelmente próxima da região Nordeste, em 2000, encontra-se muito mais próxima das regiões Sul e Sudeste em termos de desigualdade educacional. Enquanto as regiões Sudeste e Centro-Oeste, obtiveram melhoras na distribuição educacional que as aproximaram da região Sul, a região Nordeste manteve-se afastada das demais regiões permanecendo com a maior desigualdade educacional. É possível diferenciar a dinâmica do Sul e Sudeste das demais regiões, e também notar que a desigualdade educacional nestes locais sempre foi menor.

Os gráficos a seguir servem como uma base para verificação e comparação dos resultados das nossas estimativas do índice de Gini educacional. Através dessas comparações, notamos que nossas estimativas estão bastante próximas daquelas obtidas a partir de microdados, bem como a tendência de redução da desigualdade educacional, especialmente a partir da década de 1950, se confirma quando observamos cada coorte individualmente.

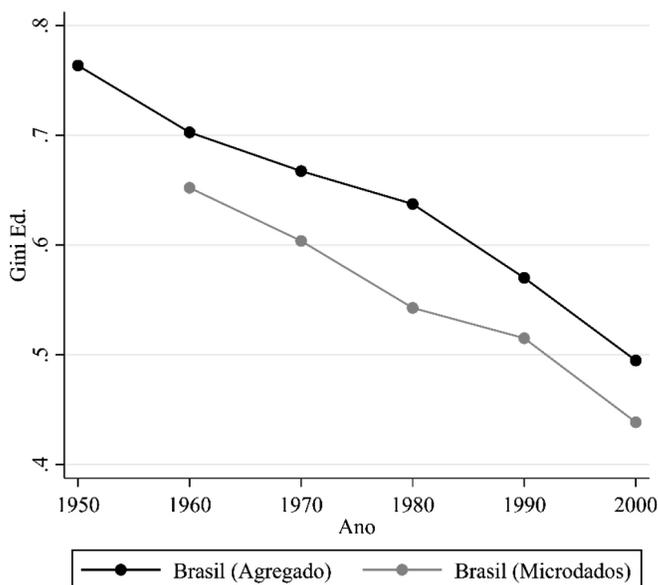


Figura 10 – Gini Educacional Microdados vs Agregados

Fonte: Censos Demográficos/IBGE. Elaboração própria.

A Figura 10 apresenta a comparação entre o cálculo do índice de Gini usando a metodologia de agregação por níveis educacionais (i.e. sem variação intragrupo) e o cálculo utilizando dados em nível individual. Ambas as séries fazem uso dos dados do Censo Demográfico, porém, apenas a partir de 1960 é possível utilizar microdados. A estimação com dados agregados situa-se acima daquela de nível individual em todos os anos. Como para definir os 5 grupos de nível de ensino foi considerado apenas o último nível concluído por cada indivíduo, ao desconsiderar a variação intragrupos, a agregação provavelmente subestima os anos de estudo da parcela educada da população e superestima o índice. No entanto, é possível observar que as tendências de longo prazo das duas séries são muito semelhantes, o que nos dá maior confiança sobre as séries estimadas para a primeira metade do século XX.

É possível perceber, entretanto, que a diferença é menor nos anos de 1960 e 1970. Dado que a proporção de pessoas com educação formal aumenta ao longo do tempo, a diferença entre as duas séries tende a aumentar com o passar do tempo. Assim, a superestimação do índice tende a ser menos expressiva do começo à metade do século, período de interesse dessa análise.

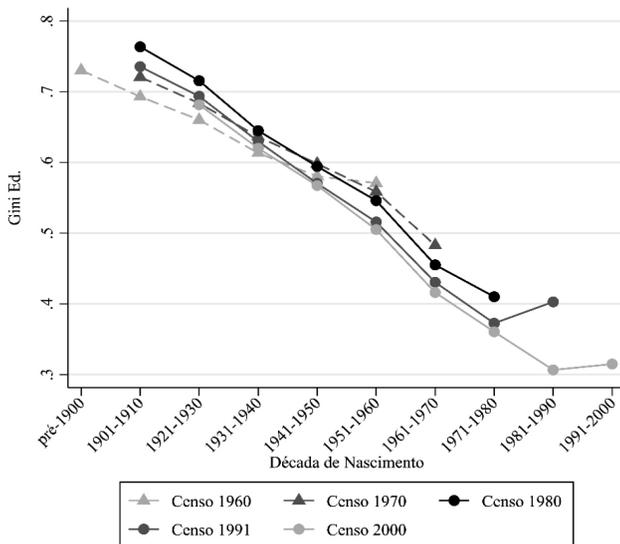


Figura 11 – Gini Educacional por Coortes de Nascimento

Fonte: Dados dos Censos Demográficos (IBGE 1960-2000). Cada uma das linhas representa um Censo Demográfico, e os pontos representam o Gini Educacional calculado para aquela coorte naquele Censo Demográfico.

A Figura 11 mostra uma forma alternativa de se estimar a evolução do índice de Gini educacional, separando os dados por coorte de nascimento de 1900 a 1990, observados nos microdados dos Censos a partir de 1960. Podemos observar que os valores do índice se reduzem continuamente a cada nova coorte, o que contrasta com a estabilidade inicial do índice de Gini educacional para todo o estoque populacional até 1940 (como visto na Figura 8). Essas diferenças podem ser explicadas por dois fatores. Em primeiro lugar, com a expansão do sistema de ensino ao longo do tempo, a tendência de redução da desigualdade educacional significa que a cada geração, mais indivíduos foram incorporados ao sistema educacional, de modo a torna-los mais homogêneos entre si em termos de escolaridade. Apesar dessas mudanças nas coortes de nascimento, as transformações ocorridas para o estoque populacional são mais lentas, uma vez que, a cada instante, há várias coortes coexistindo, e cada coorte corresponde a apenas uma pequena parcela da população total, com diferentes níveis de desigualdade educacional.

O segundo fator seria uma seleção possivelmente diferenciada por escolaridade ao longo do tempo (devido às diferenças das taxas de mortalidade), de forma que somente os mais escolarizados sobrevivem até as idades mais avançadas e são abrangidos pelos Censos a partir de 1960. No entanto, é possível notar que as linhas não se distanciam muito umas das outras, o que significa que essa seleção é pequena em comparação com a variação entre as coortes.

Dessa forma, nossa metodologia permite chegar a estimativas mais adequadas, uma vez que com ela conseguimos estimar o índice de Gini para todas as coortes em um dado instante, e não incorremos no viés de seleção devido à diferenciação das taxas de mortalidade por nível de escolaridade.

Por fim, realizamos uma comparação dos nossos resultados com aquele de Van Leeuwen e Van Leeuwen-Li (2015) para o Brasil, mostrado na Figura 12. A partir de um ponto inicial muito semelhante em 1900, fica claro que, em contraste com nossa estimativa, o Gini educacional calculado pelos autores decresce sistematicamente desde o início do século XX. A partir de 1950, quando já é possível realizar diretamente estimativas de estoques educacionais, as tendências das duas séries se igualam em um movimento decrescente, porém em níveis diferenciados. Conforme já argumentamos, no início do século, o estoque que fazia maior diferença para a desigualdade educacional era de analfabetos.

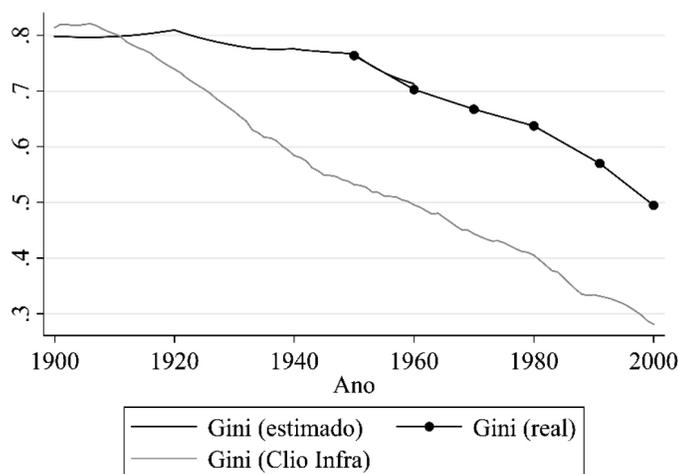


Figura 12 – Comparação de Gini Educacional por Coortes de Nascimento

Fonte: Van Leeuwen e Van Leeuwen-Li (2015); Censos Demográficos; AEB. Elaboração própria. A linha para o período entre 1900 e 1960 representam as estimativas do Gini educacional feitas a partir da composição educacional anteriormente estimada. Os pontos entre 1950 e 2000 representam o Gini educacional calculado a partir dos dados dos Censos Demográficos (IBGE 1950-2000).

6. Conclusões

Este artigo tem o intuito de contribuir para o debate acerca dos determinantes da elevada desigualdade social observada no país. Apesar da recente tendência de queda da desigualdade de renda, o Brasil ainda é um dos países mais desiguais do mundo, e parte dessa situação parece ser resultado do processo diferenciado de expansão educacional nas regiões brasileiras. Além disso, este artigo visa contribuir para a literatura apresentando estimativas mais precisas de estoques educacionais para a população brasileira construídas a partir de diversas fontes de dados históricos não utilizados com esta finalidade anteriormente. Para verificar essa hipótese, construímos séries de um índice de Gini educacional desde o início do século XX para o Brasil e as regiões analisadas.

Considerando o desafio de não haver microdados de recenseamentos populacionais disponíveis para a primeira metade do século XX, o artigo traz duas contribuições empíricas principais. Em primeiro lugar, realizamos estimativas dos estoques populacionais por escolaridade estimados para o Brasil, a partir de 1900. Na primeira metade do século XX, a maioria da população é de analfabetos, porém, o crescimento do número de pessoas com o primário completo apresenta crescimento a partir de 1940. Entre 1900 e 1930, contudo, a proporção de pessoas com ensino primário completo na população ficou em torno de 5%, ao passo que a proporção com ensino secundário completo permaneceu sempre abaixo de 1% e a parcela com ensino superior esteve sempre abaixo de 0,3%.

Nossa segunda contribuição empírica foi o cálculo de índices de Gini educacionais para o Brasil e as regiões brasileiras no mesmo período. Nossos resultados indicam que a desigualdade educacional era bastante elevada em 1900. De acordo com as nossas estimativas, naquele ano a grande maioria da população brasileira com mais de 5 anos era composta por analfabetos (68,7%) ou por pessoas sem o primário completo (25%), o que gera um índice de Gini de 0,78. O sistema educacional, em geral, pouco acessível à população, também contribuiu para a manutenção da desigualdade, de modo que a geração nascida entre 1900 e 1910 manteve uma desigualdade relativamente elevada (com índice de Gini de 0,69 em 1960).

Entre 1900 e 1920, a desigualdade educacional brasileira aumentou: consequência de um sistema educacional em expansão, mas ainda bastante restrito. Ao longo das décadas, no entanto, as novas gerações tornaram-se cada vez menos desiguais, o que é medido pelo declínio sistemático do Gini educacional. Para os nascidos na década de 1980, o índice é próximo de 0,3, tendo como referência o censo de 2000. Esse movimento foi concretizado pelo aumento de matrículas inicialmente no ensino primário, e posteriormente no secundário e superior, uma vez que no começo do século a grande maioria da população tinha escolaridade muito próxima de zero. Dessa forma, a redução da desigualdade educacional coincide com um aumento do nível médio de escolaridade. A introdução sucessiva das novas gerações mais educadas no estoque foi reduzindo a disparidade geral de educação, resultando em um movimento de redução da desigualdade no total da população entre 1920 e 2000.

No início do século XX, a desigualdade educacional era semelhante nas regiões brasileiras, com um índice de Gini entre 0,73 e 0,81. Ao longo do século, a ordenação dessas regiões em termos de desigualdade educacional não mudou, mas a diferença entre as regiões aumentou sensivelmente, em 2000, as regiões com maior e menor desigualdade educacional, a região Nordeste e a região Sudeste, tinham índices de 0,59 e 0,43 respectivamente

O acesso ao sistema de ensino e a distribuição de educação não somente são reflexos de um processo histórico de desenvolvimento, mas têm influência expressiva sobre esse mesmo processo. A região nordeste apresenta no fim do século a pior distribuição de educação, e é também aquela com pior distribuição de renda e PIB *per capita*.¹¹ Em contrapartida, a região sul, que durante a maior parte da série analisada permaneceu com a menor desigualdade de educação, no final do século figura com as primeiras colocações em distribuição de renda e PIB *per capita*.¹² Distinguir esses efeitos é uma tarefa que vai além do escopo deste trabalho, mas os resultados sugerem que há relação entre esses indicadores.

Essa evolução apresentada da desigualdade escolar ajuda a esclarecer, em parte, a desigualdade social entre as regiões. O aumento da escolaridade, associado à redução da desigualdade, parece ter um papel relevante para explicar as diferenças de renda observadas ao longo do século XX e que ainda persistem atualmente.

Referências

- Acemoglu, D., Johnson, S., e Robinson, J. A. 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Association*, 91(5): 1369-1401.
- Anazawa, L. S. 2011. *Análise da Queda na Participação no Ensino Superior*. Universidade de São Paulo: Faculdade de Economia, Administração e Contabilidade. Mimeo.
- Barro, R. J.; Sala-i-Martin; X. 1992. Convergence. *Journal of Political Economy*, 100(2): 223-251.
- Barros, A. R. 2011. *Desigualdades Regionais no Brasil: Natureza, causas, origens e soluções*. Rio de Janeiro: Elsevier.

¹¹ Instituto de Pesquisa Econômica Aplicada (IPEA). <http://www.ipeadata.gov.br/> Consultado em 21 de setembro de 2016.

¹² Instituto de Pesquisa Econômica Aplicada (IPEA). <http://www.ipeadata.gov.br/> Consultado em 21 de setembro de 2016.

- Camps, Enriqueta, e Stanley L. Engerman. 2014. The Impact of Race and Inequality on Human Capital Formation in Latin America during the Nineteenth and Twentieth Centuries. Department of Economics and Business, Universitat Pompeu Fabra, Economics Working Papers 1436.
- Central Intelligence Agency (CIA). 2013. The World Factbook 2013-14. Washington, DC: Central Intelligence Agency.
- Colistete, R. P. 2016. O Atraso em Meio à Riqueza: uma história econômica da educação primária em São Paulo, 1835 a 1920. Mimeo.
- Cowell, F. A. 2009. Measuring Inequality. LSE Perspectives in Economic Analysis. Oxford, UK: Oxford University Press.
- Cutler, D. e Deaton, A.; Lleras-Muney, A. 2006. The Determinants of Mortality. *Journal of Economic Perspectives*, 20(3): 97-120.
- Dabla-Norris, Era, Kalpana Kochhar, Frantisek Ricka, Nujin Suphaphiphat e Evridik Tsounta. 2015. Causes and Consequences of Income Inequality: A Global Perspective. International Monetary Fund, Staff Discussion Note.
- De Ferrenti, D., Perry, G. E., Ferreira, F. H. G. e Walton, M. 2004. Inequality in Latin America: Breaking with History? World Bank Latin American and Caribbean Studies; Washington, DC: World Bank.
- Directoria Geral de Estatística. 1875. Relatório Anexo ao do Minitério dos Negocios do Imperio, de 1875. Rio de Janeiro: Typographia de Pinto, Brandão e Comp.
- Diretoria Geral de Estatística (DGE). Anuário Estatístico do Brasil. Rio de Janeiro: IBGE, 1912-1936.
- Ferreira, F. H. G. 2000. Os determinantes da desigualdade de renda no Brasil: Luta de classes ou heterogeneidade educacional? In *Desigualdade e pobreza no Brasil*, organizado por Ricardo Henriques, 131-158, Rio de Janeiro: IPEA.
- Földvári, P., e Van Leeuwen, B. 2014. Educational and Income Inequality in Europe, ca. 1870-2000. *Cliometrica*, 8(3): 271-300.
- Frankema, E. 2009. The Expansion of Mass Education in Twentieth Century Latin America: a Global Comparative Perspective. *Journal of Iberian and Latin American Economic History*, 27(3): 359-395.
- Fundação Sistema Estadual de Análise de Dados (SEADE). 2010. Memória das Estatísticas Demográficas. Disponível em: <<http://produtos.seade.gov.br/produtos/500anos/index.php?tip=pri>>. Acesso em julho de 2016.
- Guizzardi Filho, O., Silva, Z. P. da, e Sidney, I. E. P. 2003. Anuários estatísticos: retratos de diferentes épocas. *São Paulo em Perspectiva*, 17(3-4): 45-54.
- Instituto Brasileiro de Geografia e Estatística (IBGE). 1938-1962. Anuário Estatístico do Brasil. Rio de Janeiro: IBGE.
- _____. 2006. Estatísticas do Século XX. Rio de Janeiro: IBGE.
- _____. 2012. Trabalho e Rendimento Educação e Deslocamento. Rio de Janeiro: IBGE.
- Instituto Nacional de Estatística. 1936-1937. Anuário Estatístico do Brasil. Rio de Janeiro: Tip do Departamento de Estatística de Publicidade.
- Kang, T. H. 2010. Instituições, Voz Política e Atraso Educacional no Brasil 1930-1964. Dissertação de Mestrado em Economia, IPE-USP, São Paulo.
- _____. 2018. Education and Development Projects in Brazil, 1932-2004: a critique. *Brazilian Journal of Political Economy*, 38(4): 766-780.
- Lam, D., e Levison, D. 1991. Declining Inequality in Schooling in Brazil and its Effects on Inequality in Earnings. *Journal of Development Economics*, 37(1-2): 199-225.
- _____, Leibbrandt, M., e Finn, A. 2015. Schooling Inequality, Returns to Schooling, and Earnings Inequality: Evidence from Brazil and South Africa. 10th IZA/World Bank Conference on Employment and Development, Bonn, Germany.
- Langoni, C. G. 1973. A Distribuição de Renda e Desenvolvimento Econômico no Brasil. Rio de Janeiro: Editora Expressão e Cultura.

- Lee, J.-W., e Lee, H. 2016. Human Capital in the Long Run. *Journal of Development Economics*, 122: 147-169.
- Leff, N. H. 1972. Desenvolvimento Econômico e Desigualdade Regional: origens do caso brasileiro. *Revista Brasileira de Economia*, 26(1): 3-21.
- Loel, B. 2008. Assessing Brazilian Educational Inequalities. *Revista Brasileira de Economia*, 62(1): 32-65.
- Lustig, N, Lopez-Calva, L. F., e Ortiz-Juarez, E. 2013. Declining Inequality in Latin America in the 2000s: The Cases of Argentina, Brazil, and Mexico. *World Development*, 44: 129-141.
- Marcílio, M. L. 2005. *História da Escola em São Paulo e no Brasil*. São Paulo: Imprensa Oficial do Estado de São Paulo, Instituto Fernand Braudel.
- Ministério da Agricultura, Industria e Commercio. Directoria Geral de Estatistica. 1916. *Anuario Estatistico do Brazil – 1º Anno (1908-1912)*. Rio de Janeiro: Typographia, vol. 1.
- Musacchio, A., Martinez, A., e Viarengo, M. 2014. Colonial Institutions, Commodity Boom, and the Diffusion of Elementary Education in Brazil, 1889-1930. National Bureau of Economic Research, NBER Working Paper 20029. Disponível em: <<http://www.nber.org/papers/w20029>>. Acesso em maio de 2016.
- Reis, J. G. A, e Barros, R. P. de. 1991. Wage Inequality and the Distribution of Education: a study of the evolution of the regional differences in inequality metropolitan Brazil. *Journal of Development Economics*, 36(1): 117-143.
- Rocha, R., Ferraz, C., e Soares, R. R. 2015. Human Capital Persistence and Development. The Institute for the Study of Labor, Discussion Paper 9101. Disponível em: <<http://ftp.iza.org/dp9101.pdf>>. Acesso em maio de 2016.
- Secretaria D’Estado dos Negocios do Imperio. 3ª Directoria. Secção de Estatística. 1887. *Trabalhos da Secção de Estatística Anexa a 3ª Directoria da Secretaria d’Estado dos Negocios do Imperio, Anno de 1886*. Rio de Janeiro: Imprensa Nacional.
- Souto, R. D. 2012. O Anuário Estatístico do Brasil e as informações sobre educação e saúde. *Estatística e Sociedade, Porto Alegre*, 2: 157-174.
- Stoops, N. S. 2006. *A Half-Century of Learning: Historical Census Statistics on Educational Attainment in the United States, 1940 to 2000*. U. S. Census Bureau, Population Division. Disponível em: <<https://www.census.gov/hhes/socdemo/education/data/census/half-century/files/US.pdf>>. Acesso em: abr. 2016.
- Thomas, V., Wang, Y., e Fan, X. 2001. *Measuring education inequality: Gini coefficients of education*. World Bank Publications.
- Tsounta, E., e Osueke, A. I. 2014. What is Behind Latin America’s Declining Income Inequality? IMF Working Paper WP/14/124.
- Van Leeuwen, B., e Leeuwen-Li, J. V. 2015. Educational Inequality Gini Coefficient, <http://hdl.handle.net/10622/KORKQW>, IISH Dataverse, v. 1.
- _____, Leeuwen-Li, J. V., e Földvari, P. 2013. Average Years of Education (Average, total Population 15 years and older), 1850-2010. Disponível em: <<https://www.clio-infra.eu/>>. Acesso em julho de 2016.
- Williamson, J. G. 2015. Latin America Inequality: Colonial Origins, Commodity Booms, or a Missed 20th Century Leveling? National Bureau of Economic Research (NBER Working Paper no. 20915). Disponível em: <<http://www.nber.org/papers/w20915>>. Acesso em maio de 2016.

Apêndice - Detalhamento sobre Dados de Ensino Superior

Exceto para três escolas de nível superior, não conseguimos encontrar dados de matrículas nesse nível de ensino para o século XIX. Na ausência dessas informações, procedemos de duas formas diferentes (que, no entanto, levaram a resultados semelhantes). Em primeiro lugar, os AEBs fornecem dados de matrículas no ensino superior em 1907 (AEB de 1953) por estado e, além disso, o número de escolas de nível superior no final do século XIX é conhecido (Anísio Teixeira 1989). Com essas duas informações, tentamos aproximar o número de matriculados existentes em 1890, supondo que em 1907 as escolas possuíam número semelhante de alunos entre si. Dessa forma, ajustamos o número de alunos de 1907 às escolas existentes em 1890, supondo adicionalmente que não houve crescimento relevante do número de vagas por instituição. Apesar de se tratar de uma aproximação, a comparação com os dados para as três escolas disponíveis mostra que os números calculados não estão muito longe da realidade. A partir desses números, calculamos o número de matrículas como proporções da população entre 20 e 24 anos.

Uma crítica que poderia ser feita ao método anterior seria de que não temos estimativas de migração dos graduados entre os estados, de modo que só conseguiríamos realizar estimativas nos estados com alguma escola de nível superior. Não é possível conhecer *a priori* a direção do viés devido a esse método em todos os estados, uma vez que é possível que estados sem escolas de nível superior tenham graduados que migraram para lá, e estados com escolas de nível superior também pode atrair graduados de outros estados. Na tentativa de lidar com esse problema, tentamos obter uma estimativa do estoque de pessoas com nível superior a partir das informações de profissão do Censo de 1872. Para isso, consideramos que todas as pessoas que exerciam algumas profissões liberais eram formadas no nível superior. A partir dessas informações, obtivemos a proporção de pessoas com ensino superior na população.

No caso da estimativa de formados no ensino superior a partir do Censo de 1872, nós aplicamos a proporção obtida à população com 25 anos ou mais em 1900 e, em seguida, aplicamos as demais proporções ao restante da população sem o ensino superior.